SUPPLEMENTARY NOTE

DETAILED DESCRIPTION OF PIQED ANALYSIS WORKFLOW

Files from DIA are processed using our custom, open-source software PIQED, which provides a graphical user interface acting as a wrapper for all the computational steps of combined discovery and quantification of PTM peptides from DIA. The GitHub repository contains detailed installation and usage instructions (https://github.com/jgmeyerucsd/PIQEDia/blob/master/Tutorial_and_manual.pdf). Data from either SCIEX or Thermo Fisher Scientific instruments are currently supported by PIQED. Raw instrument files are first converted to .mzXML using msconvert.exe from ProteoWizard, and the DIA-Umpire signal extraction module is used to generate pseudo-MS/MS spectra. MGF files produced from DIA-Umpire are automatically converted to .mzXML, and the user can choose to specify parameters for database searches using any combination of MSGF+[1], Comet[2], and/or X! Tandem[3] (see supplemental files "comet.Kac.DIA.params" and "xTandem_Kac_params.xml"). The database search results from X! Tandem and MSGF+ are automatically converted to pep.xml, whereas the comet search output should be specified as .pep.xml in the Comet configuration file. All search outputs are processed through PeptideProphet[4] using the TPP command xinteract.exe separately, combined using iProphet[5], and refined for PTM localization using PTMProphet[6]. Peptide identifications are imported into a pre-setup Skyline[7] document template using SkylineRunner.exe to generate a spectral library, perform fragment ion-level signal extraction, and output a custom fragment ion-level peak area report. The Skyline report is read into R and filtered for only peptides containing the modification of interest that have a user-defined PTMProphet localization score, optionally corrected for changes in protein level if the directory contains the file "proteinlevels.txt", and reformatted for mapDIA. The optional protein-level correction uses protein-level intensity measurements from the same sample computed using unmodified peptides (either from a separate acquisition, or in the case of the Urine data, from unmodified peptides identified

in the same acquisition) and divides all fragment-level areas for each modification site by their protein-level intensities. Optional protein-level correction will automatically be attempted by the software if the directory contains the file "proteinlevels.txt", an example of which is available on GitHub under the inputs folder. Only modification sites from proteins found in the separate protein-level quantification are included in the output, which is why the number of phosphorylation sites quantified from the urine dataset decreased after protein-level correction. See figure Supplementary Figure 4a-c for an example of protein-level normalization. The R step can also rename replicates to contain group names, which automatically happens if the file "namemapping.txt" is present in the output directory (see GitHub for an example of "namemapping.txt"). Finally, PIQED runs mapDIA[8] to perform interference filtering and to assess statistical significance of the desired comparisons. This step allows the option for total ion chromatogram normalization, or local, retention-time based normalization (see Supplementary Figure 4d-e and mapDIA publication for details). In summary, our software tool PIQED provides a high-throughput automated data processing pipeline that not only can be used to identify, but also quantify posttranslational modifications, combining many individual data processing tools in an automated fashion, which can be executed through a simple GUI and interface. At the same time, by using a DIA-only data acquisition workflow, the PTM analysis can be performed using small amounts of biological material (as no extra DDA acquisitions are needed for library building), and mass spectrometric instrument time is reduced significantly. The implementation of protein level normalization for PTM quantification allows for more precise results.

DESCRIPTION OF FILES ON GITHUB: INPUTS AND PARAMS

Diaumpire_se_orbi_strict.txt – parameters file used for DIA-Umpire signal extraction from urine dataset

Diaumpire_se.params– parameters file used for DIA-Umpire signal extraction module used for the acetyl dataset including the variable window definition

20150810.mouse.cc.iRT.fasta – database file used for MS-GF+ database searches of the acetyl dataset and for populating Skyline document

20161213.human.fasta – database file used for MS-GF+ database searches of the urine dataset and for populating the Skyline document

20150810.mouse.cc.iRT_DECOY.fasta – database file used for COMET and X! Tandem database searches of the acetyl dataset

20161213.human_DECOY.fasta – database file used for the COMET and X! Tandem database searches of the urine dataset

Comet.Kac.DIA.params – Comet database search parameters file used for acetyl dataset

Comet64.params.orbi.new – Comet database search parameters file used for urine dataset

taxonomy.xml – File required for X! Tandem database searches specifying the location of the database file used for the acetyl dataset

human_taxonomy.xml – File required for X! Tandem database searches specifying the location of the database file used for the urine dataset

xTandem_Kac_params.xml – X! Tandem database search parameters file used for the acetyl dataset

xTandem_pSTY_orbi_params.xml – X! Tandem database search parameters file used for the urine dataset

Skyline\default_empty.sky – 'empty' Skyline document containing all appropriate settings for the tutorial dataset. For user data, the template document should be edited to reflect the instrument parameters used to collect your data.

Skyline\Orbi_empty.sky – 'empty' Skyline document containing example settings for data extraction from orbitrap data.

DESCRIPTION OF FILES ON GITHUB, ACETYLLYSINE DATASET RESULTS FILES: OUTPUTS\ACETYL_MOUSE_LIVER\

fullDIA.final.interact.ptm.pep.xls.xlsx – iProphet-filtered peptide identification results

fullDIA.final.interact.ptm.pep.xml – final peptide identification results in pep.xml format

fullDIA.pt99.mProph.features.csv – complete list of mProphet feature scores produced within Skyline

2016_0826_mapDIA.skyr – custom Skyline report file

2016_0826_mapDIA.csv – Skyline report before mapDIA filtering for interferences and reformatting

mapDIA_Input.txt – Filtered and reformatted Skyline report used for input to mapDIA

site_level_areas.txt – mapDIA site-level area report containing areas used for calculation of CV values in figure 1c.

mapDIA_analysis_output.txt – raw mapDIA output results with site-level fold changes and probabilities used to generate figure 1d.

DESCRIPTION OF FILES ON GITHUB, URINE PHOSPHORYLATION DATASET RESULTS FILES: OUTPUTS\URINE\

ptmProphet-output-file.ptm.pep.xml.zip – compressed PTM prophet output

iPro-output-file.pep.xml – combined iProphet results from used for input to PTMProphet

noCor_analysis_output.txt – site-level mapDIA output using no normalization used to produce supplemental figure 3A.

proteinlevels.txt – protein-level quantities used for protein-level correction of site-level changes used to produce supplemental figure 3B.

protlvlCor_noTICcor_analysis_output.txt - site-level mapDIA output using protein-level normalization but not local TIC normalization used to produce supplemental figure 3C.

TICcor_noProtCor_analysis_output.txt - site-level mapDIA output using local TIC normalization but not protein-level normalization used to produce supplemental figure 3D.

bothCor_analysis_output.txt – site-level mapDIA output using both local TIC normalization and protein-level correction used to produce supplemental figure 3E.

SUPPLEMENTARY METHODS

SAMPLE PREPARATION – ENRICHMENT OF ACETYLATED PEPTIDES FROM MOUSE LIVER

Animal studies were performed according to protocols approved by IACUC (the Institutional Animal Care and Use Committee). Sirt5-/- male mice on 129 background were maintained on a standard chow diet (5053 PicoLab diet, Purina) until they were sacrificed at 24 weeks of age for experiments. Liver tissue from a Sirt5 knockout mouse was lysed and 20 mg of protein lysate was digested as described previously[9]. Acetylated peptides were enriched using 1 tube of anti-acetyl lysine antibody-bead conjugated PTMScan (Cell Signaling Technologies) according to the manufacturer's instructions. Peptides eluted from the enrichment were desalted using C18 reversed phase StageTips, and resuspended in 0.2% formic acid for mass spectrometry analysis.

NANO-LIQUID CHROMATOGRAPHY - TANDEM MASS SPECTROMETRY

Peptide separations were carried out using mobile phase A consisting of 97.95% water/0.05% FA/2% acetonitrile, and mobile phase B consisting of 98% acetonitrile/1.95% water/0.05% FA. Samples were loaded onto a C18 pre-column chip (200 µm x 6 mm ChromXP C18-CL chip, 3 µm, 300 Å) using an Eksigent cHiPLC system for 10 minutes at a flow of 2 µL per min of mobile phase A. Separation was performed with a 75 µm x 15 cm ChromXP C18-CL analytical chip (3 µm, 300 Å) using a gradient from 95% to 60% mobile phase A over 80 minutes. The column chip was washed by an increase to 80% B over 5 minutes that was maintained for 8 minutes, followed by a return to 95% mobile phase A over two minutes that was maintained for 35 minutes to re-equilibrate the column. Eluting peptides were directly electrosprayed into an orthogonal quadrupole time-of-flight TripleTOF 5600 mass spectrometer (SCIEX) and analyzed by data-independent acquisition (SWATH). Every SWATH cycle consisted of a 250

ms precursor scan from 400-1,250 m/z followed by fragmentation of all ions between 400-1,200 m/z using 64 variable width precursor isolation windows (window definitions contained in DIA-Umpire signal extraction parameters file "diaumpire_se.params") for 42 ms each, resulting in a total cycle time of approximately 3 sec. Fragment ion spectra were collected from 100-2,000 m/z.

PIQED SETTINGS USED FOR 1X VERSUS 2X INJECTION EXPERIMENT

All parameter files used in this study are available for download from MassIVE (MassIVE ID: MSV000080189 ftp://massive.ucsd.edu/MSV000080189/raw/). The Skyline file output from PIQED has been uploaded to Panorama at https://panoramaweb.org/labkey/project/Schilling/PIQED/begin.view (use your general Panarama login to freely access the dataset). Pseudo-MS/MS spectra from DIA-Umpire were database searched using MSGF+, Comet, and X! Tandem against the UniProt database of all mouse proteins downloaded on August 10th, 2015 and reversed sequences to allow false discovery rate (FDR) estimation using the target-decoy approach. See the supplemental files for Comet and X! Tandem search parameters. Search parameters for MSGF+ were: number of tryptic termini = 2, peptide lengths from 7-40 amino acids, charge values from 1 to 6, 20 ppm precursor mass error, and no isotope error. Searches included fixed carbamidomethylation of cysteine and up to 3 variable modifications by methionine oxidation, lysine acetylation, peptide N-terminal pyroglutamate formation, and protein N-terminal acetylation. Database search results were converted from .mzid to .pep.xml using idconvert.exe, and results were refined using PeptideProphet. PeptideProphet processing used the following options: do not merge files into one analysis, use PPM mass error model, use the non-parametric model with decoy hits to pin down the negative distribution, report decoy hits with computed probability, enzyme trypsin, and clevel=0. After separate refinement with PeptideProphet, all database search results were combined into one file using iProphet with the default parameters. The

iProphet output was then used as input to PTMProphet for PTM localization scoring. Finally, the peptide identification results in .pep.xml format were imported into Skyline to generate a spectral library. SWATH fragment ion peak areas from the three replicates of 0.5X and 1X injections were exported as the custom Skyline reports and peptides containing lysine + 42 (acetylation) were filtered based on a minimum localization score of 0.95. mapDIA, which filters fragment ions and uses well-established Bayesian modeling to compute the false discovery rates of changes, was then used to compare the two groups (i.e. "halfDIA" and "fullDIA") with the default parameters using no normalization. The mapDIA model is quite robust with any distribution of input data, therefore no assumptions are required for its use to test differential expression. The results presented here are representative of several similar experiments performed in our laboratory.

PIQED ANALYSIS OF URINE DATA FROM Q-EXACTIVE

Eleven instrument .raw files per diagnosis group were downloaded from the set of files available on PeptideAtlas (dataset identifier: PASS00706) published by Muntel *et al*.[10] PIQED setting for this dataset were as described above except if noted differently here. Because this data was collected without phosphopeptide enrichment, we used more stringent signal extraction and post-quantification filtering parameters than those described in the previous section. The suggested DIA-Umpire signal extraction parameters included with the DIA-Umpire download were used except that CorrThreshold was set to 0.5, DeltaApex was set to 0.1, RTOverlap was set to 0.1, BoostComplementaryIon was set to FALSE, SE.estimateBG was set to TRUE, and SE.MassDefectOffset was set to 0.2. Database searches were performed against the human proteome downloaded from UniProt on December 13th, 2016, and used 15 ppm precursor tolerance. Searches allowed variable phosphorylation of STY and variable oxidation of methionine.

STATISTICS

Acetylated peptide data used for the plots in the main text figure was from three technical replicates corresponding to repeat injections of 0.5X or 1X sample volume from the same sample. The acetylated peptide results presented in figure 1 are representative of 2 repetitions of the experiment. Phosphorylated peptide data comes from reanalysis of biological replicates corresponding to urine collected from unique patients, eleven from each diagnosis group (see supplemental reference 10 for more sample details). For statistical details of each individual software program, see their respective publications. Each of the external programs (used as part of our pipeline) contains their own thorough statistical algorithms and significance testing. No additional statistical tests were implemented outside of the constitutive programs.

CODE AVAILABILITY

The current PIQED software is available from GitHub (https://github.com/jgmeyerucsd/PIQEDia/releases/tag/v0.1.2), GNU General Public License v3.0. Updated and future software versions will be available from https://github.com/jgmeyerucsd/PIQEDia .

SUPPLEMENTARY REFERENCES

1. Kim, S. & Pevzner, P. A. MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat Commun* **5,** (2014).

2. Eng, J. K., Jahan, T. A. & Hoopmann, M. R. Comet: An open-source MS/MS sequence database search tool. *PROTEOMICS* **13,** 22–24 (2013).
3. Craig, R. & Beavis, R. C. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **20,** 1466–1467 (2004).
4. Keller, A., Nesvizhskii, A. I., Kolker, E. & Aebersold, R. Empirical Statistical Model To Estimate the Accuracy of Peptide Identifications Made by MS/MS and Database Search. *Anal. Chem.* **74,** 5383–5392 (2002).
5. Shteynberg, D. *et al.* iProphet: Multi-level Integrative Analysis of Shotgun Proteomic Data Improves Peptide and Protein Identification Rates and Error Estimates. *Mol. Cell. Proteomics* **10,** (2011).
6. Shteynberg, D. *et al.* PTMProphet: TPP software for validation of modified site locations on post-translationally modified peptides. in *60th ASMS Conference on Mass Spectrometry, Vancouver, BC, Canada* 20–24 (2012).
7. Egertson, J. D., MacLean, B., Johnson, R., Xuan, Y. & MacCoss, M. J. Multiplexed peptide analysis using data-independent acquisition and Skyline. *Nat Protoc.* **10,** 887–903 (2015).
8. Teo, G. *et al.* mapDIA: Preprocessing and statistical analysis of quantitative proteomics data from data independent acquisition mass spectrometry. *J. Proteomics* **129,** 108–120 (2015).
9. Rardin, M. J. *et al.* SIRT5 regulates the mitochondrial lysine succinylome and metabolic networks. *Cell Metab.* **18,** 920–933 (2013).
10. Muntel, J. *et al.* Advancing Urinary Protein Biomarker Discovery by Data-Independent Acquisition on a Quadrupole-Orbitrap Mass Spectrometer. *J. Proteome Res.* **14,** 4752–4762 (2015).